LETTER

doi:10.1038/nature13899

Structural insight into autoinhibition and histone H3–induced activation of DNMT3A

Xue Guo[1,2]*, Ling Wang[1,2]*, Jie Li[1], Zhanyu Ding[3], Jianxiong Xiao[1], Xiaotong Yin[1], Shuang He[1], Pan Shi[4,5,6], Liping Dong[7,8], Guohong Li[7], Changlin Tian[4,5,6], Jiawei Wang[9], Yao Cong[3] & Yanhui Xu[1,2]

DNA methylation is an important epigenetic modification that is essential for various developmental processes through regulating gene expression, genomic imprinting, and epigenetic inheritance[1–5]. Mammalian genomic DNA methylation is established during embryogenesis by *de novo* DNA methyltransferases, DNMT3A and DNMT3B[6–8], and the methylation patterns vary with developmental stages and cell types[9–12]. DNA methyltransferase 3-like protein (DNMT3L) is a catalytically inactive paralogue of DNMT3 enzymes, which stimulates the enzymatic activity of Dnmt3a[13]. Recent studies have established a connection between DNA methylation and histone modifications, and revealed a histone-guided mechanism for the establishment of DNA methylation[14]. The ATRX–DNMT3–DNMT3L (ADD) domain of Dnmt3a recognizes unmethylated histone H3 (H3K4me0)[15–17]. The histone H3 tail stimulates the enzymatic activity of Dnmt3a *in vitro*[17,18], whereas the molecular mechanism remains elusive. Here we show that DNMT3A exists in an autoinhibitory form and that the histone H3 tail stimulates its activity in a DNMT3L-independent manner. We determine the crystal structures of DNMT3A–DNMT3L (autoinhibitory form) and DNMT3A–DNMT3L-H3 (active form) complexes at 3.82 and 2.90 Å resolution, respectively. Structural and biochemical analyses indicate that the ADD domain of DNMT3A interacts with and inhibits enzymatic activity of the catalytic domain (CD) through blocking its DNA-binding affinity. Histone H3 (but not H3K4me3) disrupts ADD–CD interaction, induces a large movement of the ADD domain, and thus releases the autoinhibition of DNMT3A. The finding adds another layer of regulation of DNA methylation to ensure that the enzyme is mainly activated at proper targeting loci when unmethylated H3K4 is present, and strongly supports a negative correlation between H3K4me3 and DNA methylation across the mammalian genome[9,10,19,20]. Our study provides a new insight into an unexpected autoinhibition and histone H3-induced activation of the *de novo* DNA methyltransferase after its initial genomic positioning.

To investigate how DNMT3A activity is regulated, we performed an *in vitro* DNA methylation assay using recombinant DNMT3A2 (residues 224–912), a catalytically active variant of DNMT3A consisting of Pro-Trp-Trp-Pro (PWWP), ADD, and CD domains (Fig. 1a)[21]. H3K4me0 (but not H3K4me3) peptide significantly stimulated the enzymatic activities of DNMT3A2 or DNMT3A2 in complex with the CD-like domain of DNMT3L (designated C^{DNMT3L}) (Fig. 1b and Extended Data Fig. 1a, b). A similar effect was observed for DNMT3A2 or DNMT3A2–C^{DNMT3L} using recombinant poly-nucleosomes carrying unmethylated histone H3 or H3K4me3 mimic ($H3K_C4me3$) (Fig. 1b and Extended Data Fig. 1c). Thus, the enzymatic activity of DNMT3A2 is stimulated by histone H3 (but not H3K4me3) tail either in the form of free peptide, or within nucleosome, in a DNMT3L-independent manner.

We next characterized the critical regions for histone H3-induced activation of DNMT3A. H3K4me0 peptide activated DNMT3A2 and ADD–CD (residues 476–912) proteins with comparable (approximately sixfold) activity enhancement, but did not stimulate activity of the CD domain (residues 627–912) in the absence or presence of C^{DNMT3L} (Extended Data Fig. 1d, e). ADD–CD protein with histone H3 peptide (residues 1–20) fused at the amino (N) terminus (H3–ADD–CD) showed significantly higher activity than ADD–CD, and could not be further activated by H3K4me0 peptide (Extended Data Fig. 1e). These results collectively indicate that the ADD domain (but not the PWWP domain) is required for histone H3-induced activation of DNMT3A.

Notably, the CD domain or CD–C^{DNMT3L} complex showed comparable activity to that of ADD–CD or ADD–CD–C^{DNMT3L} activated by histone H3 peptide (Extended Data Fig. 1d, e), suggesting that the ADD domain inhibit the activity of the CD domain and histone H3 release this inhibition. In support of this hypothesis, addition of ADD–linker (residues 476–626, Fig. 1a), but not H3–ADD–linker fusion protein inhibited the activity of the CD domain (Fig. 1c). These results confirm the existence of autoinhibition of DNMT3A, in which ADD–linker directly inhibits the enzymatic activity of the CD domain, and that this inhibition is released by histone H3 tail.

We determined the crystal structure of ADD–CD of DNMT3A in complex with C^{DNMT3L} (ADD–CD–C^{DNMT3L}) at 3.82 Å resolution (Fig. 1d and Extended Data Table 1). Although only one ADD–CD–C^{DNMT3L} complex was observed within an asymmetric unit of the crystals, gel-filtration analysis indicated a tetramer formation in solution. Two ADD–CD–C^{DNMT3L} complexes form dimer of dimers via the CD–CD interaction in a two-fold crystallographic symmetry (Extended Data Fig. 2). The complex structure adopts an 'X' shape with two CD domains located in the centre and the ADD and C^{DNMT3L} domains located in the four corners. In the ADD–CD–C^{DNMT3L} structure, CD–C^{DNMT3L} adopts a similar fold to that of mouse CD^{Dnmt3a}–C^{Dnmt3L} structure (Extended Data Fig. 3a)[22].

The ADD and CD domains fold into two individual structural modules connected by the linker (Fig. 1d). The linker forms a twisted helix followed by an extended loop, which packs against a hydrophobic surface of the CD domain (Extended Data Fig. 3b, c). A loop region (residues 526–533) extends out of the ADD domain and inserts into a pocket in the CD domain (Fig. 1e and Extended Data Fig. 3d). This pocket is mainly formed by basic residues R790, R792, H789, and R831 of the CD domain. Three acidic residues (D529/D530/D531) and hydrophobic residues (Y526/Y528/Y533) of the ADD domain are brought into close proximity to the pocket. All residues that are involved in the intramolecular interaction are highly conserved in DNMT3A/3B, suggesting a conserved ADD–CD interface among DNMT3A/B sub-family

[1]Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai 200032, China. [2]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433, China. [3]National Center for Protein Science Shanghai, State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [4]High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230031, China. [5]National Laboratory for Physical Science at the Microscale, University of Science and Technology of China, Hefei 230026, China. [6]School of Life Sciences, University of Science and Technology of China, Hefei 230026, China. [7]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China. [8]University of Chinese Academy of Science, Beijing 100049, China. [9]State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.

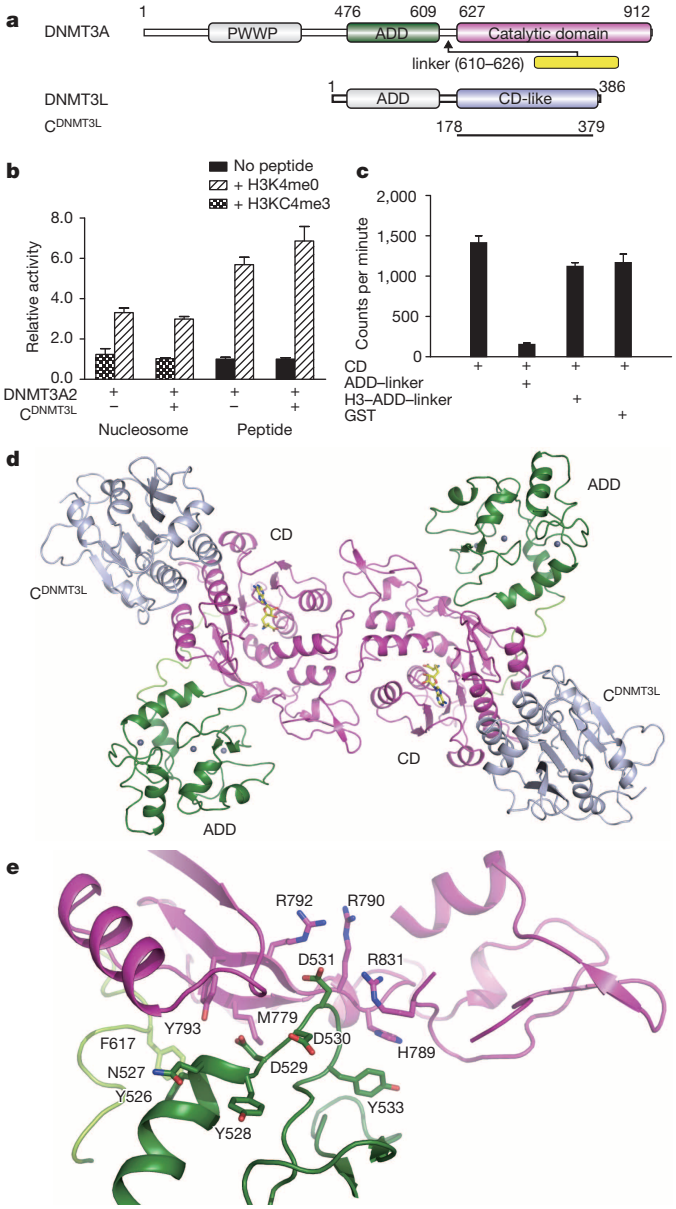


Figure 1 | Structure of the DNMT3A–DNMT3L complex in autoinhibitory form. **a**, Colour-coded domain architecture of human DNMT3A and DNMT3L. **b**, Relative activities of DNMT3A2 using peptides or reconstituted nucleosomes as substrates in the absence or presence of C^{DNMT3L}. **c**, Activities of the CD domain in the absence or presence of ADD–linker or H3–ADD–linker fusion proteins. **d**, Ribbon representations of the overall structure of the ADD–CD–C^{DNMT3L} complex. AdoHcy are shown in stick representation; zinc cations are shown as grey balls. The colour scheme is the same as in **a** and used in all structural figures. **e**, Close-up view of the ADD–CD interface. Critical residues for the interactions are shown in stick representation. Relative activities were calculated according to basal activity for each protein in the absence of H3 peptide. Error bars, s.d. for triplicate experiments.

members, but not DNMT3L (Extended Data Fig. 4). Structural comparison also indicated that ADD^{DNMT3L} would have steric hindrance with C^{DNMT3L} if DNMT3L adopts a similar conformation to that of DNMT3A (Extended Data Fig. 3e)[15].

The intramolecular interaction was verified in the *in vitro* glutathione *S*-transferase (GST) pull-down assays, in which ADD–linker bound to the CD domain (Fig. 2a and Extended Data Fig. 5a–e). Mutations D529A, D531A, D529A/D531A, and Y526A/Y528A of ADD–linker, decreased the binding affinity to the CD domain, supporting their important role in mediating the ADD–CD interaction. In contrast, mutations R556A, M548W, D530A, and Y533A of ADD–linker showed little effect on ADD–CD interaction (Fig. 2a and Extended Data Fig. 5a, b). Mutation R790A/R792A of the CD domain also impaired its binding affinity to the ADD–linker protein (Extended Data Fig. 5c). H3K4me0 (but not H3K4me3) peptide hampered ADD–CD interaction, suggesting that histone H3 tail releases autoinhibition of DNMT3A through disrupting this intramolecular interaction (Fig. 2b and Extended Data Fig. 5d).

We next mapped the critical regions for the autoinhibition of DNMT3A in the presence of C^{DNMT3L} (Fig. 2c). Compared with ADD–CD (Del-N476), Del-N522 showed a slight increase in enzymatic activity, whereas Del-N550, Del-N587, and Del-N610 significantly increased their activities (Fig. 2d). Although ADD and ADD–linker showed comparable binding affinity to the CD domain (Extended Data Fig. 5e), only ADD–linker (ADD-626) maintained the inhibitory function, while all other truncated proteins markedly decreased their inhibition on the activity of the CD domain (Fig. 2e). Replacement of residues 621–632 by a 12-residue Gly and Ser (GS) linker in ADD–CD partly released autoinhibition and showed less than twofold activity enhancement upon histone H3 stimulation (Extended Data Fig. 5f). The results indicate that the ADD domain and linker are both important for the autoinhibition of DNMT3A.

We next tested whether residues on the ADD–CD interface are important for the autoinhibition of DNMT3A. Compared with wild-type protein, mutations D529A, D531A, and Y526A/Y528A of the ADD–linker could barely inhibit the activity of the CD domain (Fig. 2f and Extended Data Table 2). Mutations D529A, D531A, and Y526A/Y528A of DNMT3A2 showed enzymatic activities that are comparable to or slightly lower than that of corresponding mutants activated by histone H3 peptide (Fig. 2g). A more obvious release of autoinhibition was observed in assays using recombinant poly-nucleosomes as substrate (Fig. 2h). As a negative control, mutation M548W of DNMT3A2 remained in an inhibitory form and was not activated by histone H3 tail because the mutant could not bind to histone H3 (ref. 18). Taken together, the ADD domain and linker function together to inhibit the activity of the CD domain, and ADD–CD association is essential for ADD-mediated autoinhibition of DNMT3A.

HhaI is a well-characterized bacterial DNA (cytosine-5) methyltransferase[23]. Structural comparison of ADD–CD–C^{DNMT3L} and HhaI–DNA (Protein Data Bank (PDB) accession number 1MHT) complexes indicates that their catalytic domains adopt similar fold, and suggests that loops L1 and L2 of DNMT3A are potential regions for DNA interaction (Fig. 3a–c and Extended Data Fig. 4). Consistently, mutating basic residues K831, R836/N838, K841, and K844 on loop L2 impaired the enzymatic activity of DNMT3A (Fig. 3c and Extended Data Fig. 6a)[24]. In the ADD–CD–C^{DNMT3L} structure, the ADD domain is located close to the loop L2 and may thus generate steric hindrance for DNA interaction (Fig. 3c).

In support of the analysis above, ADD–CD markedly decreased, but H3–ADD–CD showed comparable DNA-binding affinity to that of the CD domain (Fig. 3d). Histone H3 peptide partly restored DNA-binding affinity of ADD–CD or DNMT3A2, which may have resulted from relative weak interaction between histone H3 peptide and DNMT3A2 (Fig. 3d, e and Extended Data Fig. 6b). Consistent with their effect on the release of autoinhibition, mutations D529A, D531A, and Y526A/Y528A of DNMT3A2–C^{DNMT3L} partly rescued the DNA interaction (Fig. 3e), and histone H3 (but not H3K4me3) peptide increased the complex formation of ADD–CD–DNA, but not CD–DNA (Fig. 3f). Similar results were observed in electrophoretic mobility-shift assays (Extended Data Fig. 6c). Taken together, the ADD domain inhibits the activity of the CD domain by decreasing its DNA-binding affinity. This inhibition is restored by fusion of the histone H3 tail at the N terminus of ADD–CD, and is partly restored by addition of histone H3 peptide or mutations of residues on the ADD–CD interface.

We also determined the crystal structure of ADD–CD–C^{DNMT3L} in complex with histone H3 peptide (residues 1–12) at 2.90 Å resolution (Extended Data Table 1). The complex structure adopts a butterfly shape, with the ADD and C^{DNMT3L} domains resembling the wings (Fig. 4a). In

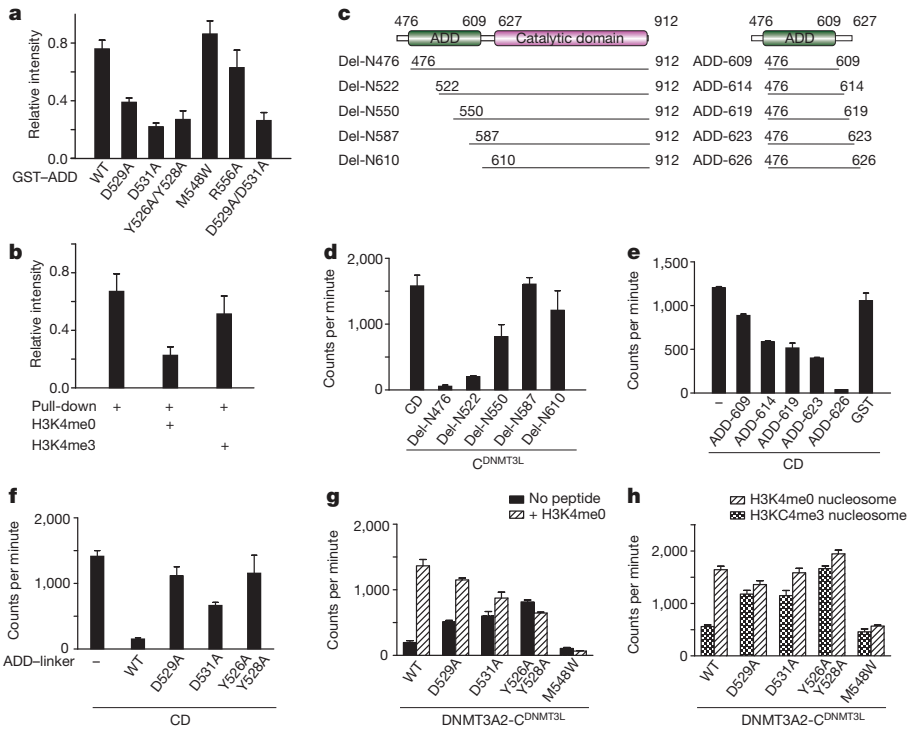


Figure 2 | ADD–CD interactions are important for autoinhibition of DNMT3A. **a**, GST pull-down assays for the ADD–CD interactions. Recombinant CD (residues 627–912) proteins were incubated with wild-type or mutant GST–ADD–linker proteins immobilized on glutathione resin. The bound proteins were analysed using SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and Coomassie blue staining (Extended Data Fig. 5a). The assays were quantified by band densitometry. Error bars, s.d. for triplicate experiments. **b**, GST pull-down assays in the absence or presence of histone H3 peptide (H3K4me0 or H3K4me3) as indicated in Extended Data Fig. 5d. **c**, Schematic representation of DNMT3A proteins used for *in vitro* methyltransferase activity assays. **d**, Effect of various N-terminal deletions of ADD–CD on its enzymatic activity. **e**, Activities of the CD domain measured in the presence of various carboxy (C)-terminal deletions of ADD–linker proteins. **f**, Activities of the CD domain measured in the presence of wild-type and mutant ADD–linker proteins. **g**, **h**, Activities of wild-type or mutant DNMT3A2 measured using naked DNA (**g**) or poly-nucleosomes (**h**) as substrate. Error bars, s.d. for triplicate experiments.

the complex structure, two ADD–CD–C^{DNMT3L}-H3 complexes form a dimer of dimers through CD–CD interface in a pseudo-two-fold symmetry. CD–CD^{DNMT3L} adopts a similar fold to that of ADD–CD–C^{DNMT3L} (autoinhibitory form) and mouse CD^{Dnmt3a}–C^{Dnmt3L} structures, whereas the position of the ADD domain is obviously different in the structures of DNMT3A in autoinhibitory and active forms (Fig. 4a, b and Extended Data Fig. 7a, b).

In the ADD–CD–C^{DNMT3L}-H3 structure, the ADD–CD interaction is mediated by a network of hydrogen bonds and hydrophobic interactions (Extended Data Fig. 7c, d). A similar domain organization was observed in the DNMT3L–H3 structure[15], suggesting a conserved ADD–CD interface among DNMT3 family members when they adopt the active form (Extended Data Fig. 7e). The histone H3 peptide binds to the ADD domain by forming a three-stranded anti-parallel β-sheet with two β-strands of the ADD domain (Extended Data Fig. 7f). The side chain of H3K4 is stabilized by residues D529 and D531 of the ADD domain, and tri-methylation of H3K4 will disrupt the interactions. Mutations D529A, D531A, and M548W of ADD–CD abolished their binding affinity to histone H3 peptide (Extended Data Fig. 7g).

Comparison of the structures of ADD–CD–C^{DNMT3L} in autoinhibitory and active forms indicates two separate surfaces on the CD domain for ADD–CD interaction, and suggests a conformational change of DNMT3A

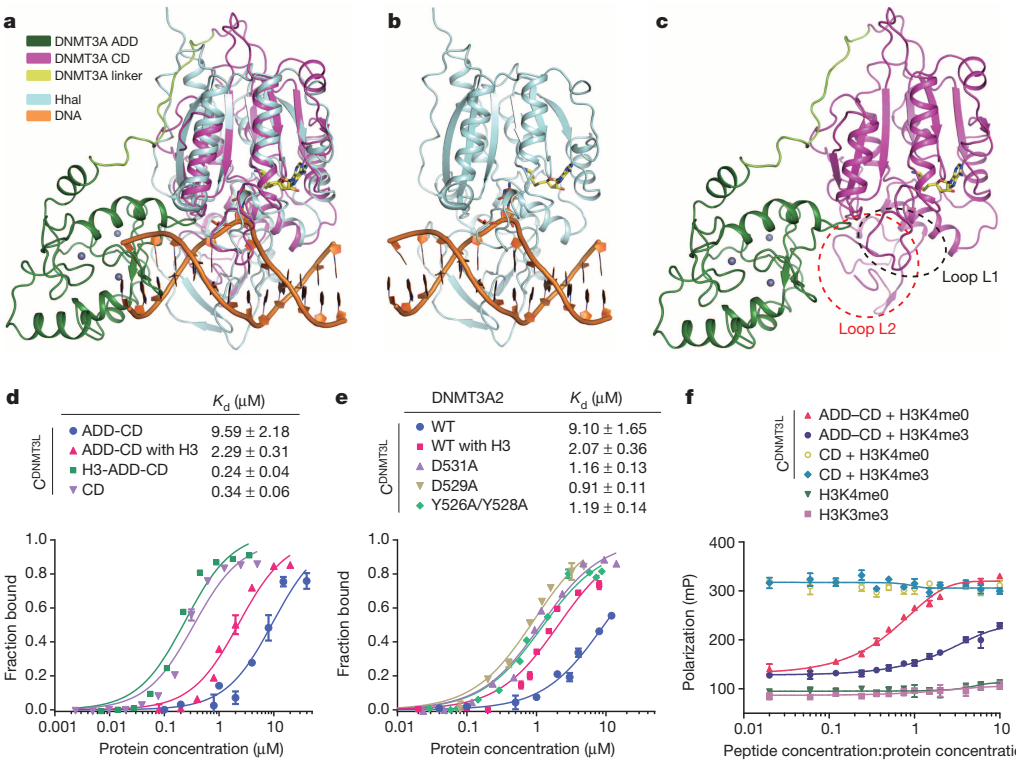


Figure 3 | The ADD domain inhibits DNA-binding affinity of the catalytic domain. **a**, Superimposition of ADD–CD–C^{DNMT3L} and HhaI–DNA complex structures shown in ribbon representations with the C^{DNMT3L} domain omitted for simplicity. The colour scheme for the comparison is indicated. **b**, **c**, HhaI–DNA (**b**) and ADD–CD (**c**) structures are shown as in **a**. **d**, **e**, Superimposed fluorescence polarization plots for DNA-binding affinities of truncations (**d**) or mutants (**e**) of DNMT3A in the absence or presence of histone H3 peptide. **f**, Superimposed fluorescence polarization plots for protein–DNA complex formation of various truncations of DNMT3A in the absence or presence of histone H3 peptides. The levels of protein–DNA complex formation are represented by fluorescence polarization response (in milli-polarization units, mP).

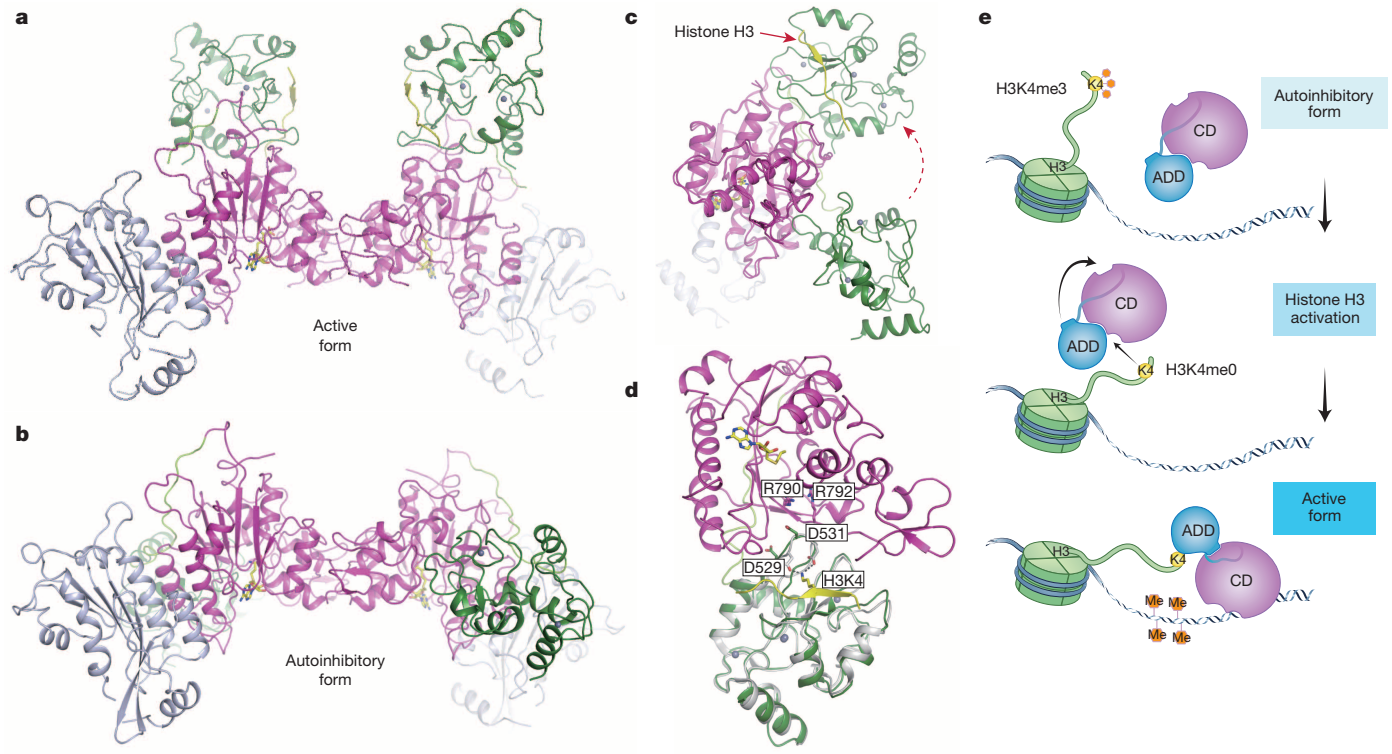


Figure 4 | Mechanisms for autoinhibition and histone H3 tail-induced activation of DNMT3A. **a**, Ribbon representations of the overall structure of the ADD–CD–C^{DNMT3L}-H3 complex. Histone H3 peptides are coloured in yellow. **b**, ADD–CD–C^{DNMT3L} structure in autoinhibitory form are shown for comparison. **c**, Ribbon representations of ADD–CD–C^{DNMT3L} (autoinhibitory form) and ADD–CD–C^{DNMT3L}-H3 (active form) complex structures with the CD domains superimposed. **d**, Superimposition of ADD–CD–C^{DNMT3L} (C^{DNMT3L} omitted for simplicity) and H3–ADD structures. The ADD domain and histone H3 in the H3–ADD structure are coloured in grey and yellow, respectively. Residues for H3–ADD and ADD–CD interactions are shown in stick representation. **e**, A working model for the autoinhibition and histone H3 tail-induced activation of DNMT3A. DNMT3L and the PWWP domain of DNMT3A are not shown for simplicity.

induced by histone H3 (Fig. 4a–c and Extended Data Fig. 8a). DNMT3A may prefer an autoinhibitory form in the absence of histone H3. Upon histone H3 association, DNMT3A adopts an active form to open an otherwise locked DNA-binding region for establishment of DNA methylation. The negative-stain electron microscopy density maps clearly showed a distinct shape for DNMT3A2 (residues 275–912)–C^{DNMT3L} complex in the absence or presence of histone H3 peptide, supporting a conformational change of DNMT3A2–C^{DNMT3L} induced by histone H3: that is, transforming from an 'X' shape (autoinhibitory form) to a butterfly shape (active form) (Extended Data Fig. 8b–d). One-dimensional ^{19}F NMR measurement[25] also indicated a significant change of chemical shift (representing conformational change) for residue F827 (within the loop L2) upon histone H3 peptide (but not H3K4me3) titration, whereas only a slight change was observed for residue F868 (as a negative control) (Extended Data Fig. 8e, f). Collectively, both electron microscopy and NMR measurements support the existence of conformational change of DNMT3A induced by histone H3 tail.

Superimposition of H3–ADD (PDB accession number 3A1B) and ADD–CD–C^{DNMT3L} (autoinhibitory form) structures indicated that ADD domains in the two structures adopt a similar fold, and that histone H3 peptide in the H3–ADD structure has no overlap with the ADD–CD interface (Fig. 4d). Our previous studies have shown that mutation D529A or D531A not only significantly decreased H3–ADD interaction (Extended Data Fig. 7g) but also markedly decreased ADD–CD interaction (Fig. 2a), and released the autoinhibition (Fig. 2g, h). Moreover, H3K4me0 peptide could disrupt ADD–CD interaction (Fig. 2b). Thus, histone H3 releases the autoinhibition of DNMT3A through binding to residues of the ADD domain on the ADD–CD interface (such as D529 and D531) and disrupting the intramolecular interaction. The two residues function as a critical switch that can exist in both forms of DNMT3A and couple histone H3 recognition to the release of the autoinhibition.

Here we propose a working model for histone H3-induced dynamic regulation of the *de novo* DNA methyltransferase (Fig. 4e and Supplementary Video 1). DNMT3A exists in dynamic equilibrium between autoinhibitory and active forms, and the ADD domain oscillates between the two conformations. In the absence of histone H3, DNMT3A prefers an autoinhibitory form, in which the ADD domain binds to the CD domain and hinders its DNA-binding affinity. Once DNMT3A (or the DNMT3A–DNMT3L complex) is recruited to the nucleosome, H3K4me0 binds to the ADD domain and stimulates DNMT3A to undergo a significant conformational change from an autoinhibitory form to an active form. The ADD–CD interaction in the active form allows DNMT3A to adopt a relative stable conformation so that DNA methylation occurs in a range permitted by the histone H3 tail. Even when DNMT3A is recruited to an H3K4me3-containing nucleosome, the enzyme will remain in its autoinhibitory form to avoid DNA methylation within an impermissible chromatin environment. Our work reinforces the connection between DNA methylation and histone modifications, and sheds new light on the fine tuning of the establishment of DNA methylation. The structures may also provide a basis for the design of specific regulators for potential therapeutic applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 May; accepted 29 September 2014.
Published online 10 November 2014.

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (suppl.), 245–254 (2003).
2. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14,** 204–220 (2013).

3. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11,** 204–220 (2010).
4. Guibert, S., Forne, T. & Weber, M. Dynamic regulation of DNA methylation during mammalian development. *Epigenomics* **1,** 81–98 (2009).
5. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
6. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genet.* **19,** 219–220 (1998).
7. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
8. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9,** 2395–2402 (2000).
9. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20,** 320–331 (2010).
10. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
11. Bird, A. P. & Taggart, M. H. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res.* **8,** 1485–1497 (1980).
12. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **10,** 2709–2721 (1982).
13. Chedin, F., Lieber, M. R. & Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl Acad. Sci. USA* **99,** 16916–16921 (2002).
14. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Rev. Genet.* **10,** 295–304 (2009).
15. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
16. Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* **10,** 1235–1241 (2009).
17. Zhang, Y. *et al.* Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* **38,** 4246–4253 (2010).
18. Li, B. Z. *et al.* Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Res.* **21,** 1172–1181 (2011).
19. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39,** 457–466 (2007).
20. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
21. Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J. Biol. Chem.* **277,** 38746–38754 (2002).
22. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* **449,** 248–251 (2007).
23. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. Hhal methyltransferase flips its target base out of the DNA helix. *Cell* **76,** 357–369 (1994).
24. Rajavelu, A., Jurkowska, R. Z., Fritz, J. & Jeltsch, A. Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic Acids Res.* **40,** 569–580 (2012).
25. Shi, P. *et al.* Site-specific protein backbone and side-chain NMR chemical shift and relaxation analysis of human vinexin SH3 domain using a genetically encoded 15N/19F-labeled unnatural amino acid. *Biochem. Biophys. Res. Commun.* **402,** 461–466 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank staff of beamline BL17U at Shanghai Synchrotron Radiation Facility, China, for their assistance in data collection, and H. Wang for help on electron microscopy analyses. We thank staff of the Biomedical Core Facility, Fudan University, for their help on biochemical analyses, and A. D. Riggs for providing the complementary DNAs of DNMT3A and DNMT3L. This work was supported by grants from the National Basic Research Program of China (2011CB965300, 2009CB918600, 2013CB910401), the National Science & Technology Major Project 'Key New Drug Creation and Manufacturing Program' of China (2014ZX09507-002, 2011ZX09506-001), the National Natural Science Foundation of China (31270779, 91419301, 31030019, U1432242, 31270771, 31222016, 31300685, U1332138), the Basic Research Project of Shanghai Science and Technology Commission (12JC1402700, 13JC1406300), the Fok Ying Tung Education Foundation (20090071220012), and the Chinese Academy of Sciences Pilot Strategic Science and Technology Projects B (numbers XDB08030201, XDB08030302). Y.C. is a scholar of the Hundred Talents Program of the Chinese Academy of Sciences.

Author Contributions X.G., L.W., and Y.X. designed the experiments. X.G., L.W., and X.Y. performed protein purification and crystallization of ADD–CD–C^{DNMT3L}-H3 complex; X.G. collected the data and determined the crystal structure. L.W., J.X., and S.H. performed protein purification, crystallization, and data collection of ADD–CD–C^{DNMT3L}. J.L. and J.W. determined the crystal structure. X.G. and L.W. performed enzymatic assays, fluorescence polarization and pull-down assays. L.D. and G.L. prepared nucleosomes for assays. C.T. and P.S. performed and analysed the ^{19}F NMR measurements. Z.D. and Y.C. performed and analysed the electron microscopy measurements. X.G., L.W., and Y.X. analysed the data and wrote the manuscript. Y.X. supervised the project.

Author Information The coordinates and structure factors for the ADD–CD–C^{DNMT3L} and ADD–CD–C^{DNMT3L}-H3 structures have been deposited in the Protein Data Bank under accession numbers 4U7P and 4U7T, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.X. (xuyh@fudan.edu.cn).

METHODS

Protein expression and purification. The full-length DNMT3A2 (residues 224–912 of DNMT3A) was expressed in sf9 cells using the Bac-to-Bac system (Invitrogen). The infected cells were harvested and lysed in 50 mM Tris-Cl pH 8.0, 500 mM NaCl, and 0.01% 2-mercaptoethanol. The clarified lysate was applied onto GST affinity columns (GE Healthcare) and the fusion protein was cleaved with PreScission protease. The protein was stored in 20 mM Tris-Cl pH 8.0, 300 mM NaCl, and 0.01% 2-mercaptoethanol for the assays.

Truncations of DNMT3A2 were cloned into modified pGEX-6p-1 vector and the C^{DNMT3L} (residues 178–379) was inserted into modified pRSFDuet-1 vector. DNMT3A proteins were expressed independently or with C^{DNMT3L} in *Escherichia coli* strain BL21(DE3). The transformants were grown at 37 °C to an attenuance ($D_{600\,nm}$) of 0.6 in 2× YT medium. The cultures were induced by adding 1 mM isopropyl-β-D-thiogalactopyranoside and further incubated for 16 h at 15 °C. The supernatant of cell lysate was applied onto GST affinity columns (GE Healthcare) and the fusion protein was digested with PreScission protease. The eluted protein was purified by ion exchange and gel filtration chromatography. The purified proteins were subjected to SDS–PAGE and stained by Coomassie blue. The peak fractions were concentrated to 5–10 mg ml^{-1} and used for crystallization and biochemical assays. Mutations of DNMT3A were purified in a similar procedure.

Crystallization and structure determination. For crystallization of the autoinhibitory form of DNMT3A, the complex of ADD–CD of DNMT3A (residues 455–912) and C^{DNMT3L} was mixed with a palindromic 18-base-pair DNA duplex (5′-GAGGCTAGCGCTAGCCTC-3′) and AdoHcy in a 1:1.2:2 molar ratio. The crystals were obtained using the hanging-drop, vapour-diffusion method by mixing 1 μl ADD–CD–C^{DNMT3L} complex with 1 μl reservoir solution containing 0.05 M Bis-Tris pH 5.6–6.0, 0.1 M sodium malonate and 8% PEG3350 at 4 °C. Crystals were cryoprotected by the reservoir buffer containing 22% glycerol and then flash frozen in liquid nitrogen. Although the DNA duplex was added for the crystallization, no corresponding electron density was observed. The DNA duplex may function as an additive to favour the crystallization.

For the crystallization of the active form of DNMT3A, the complex of ADD–CD of DNMT3A (residues 476–912) and C^{DNMT3L} was pre-incubated with histone H3 peptide (residues 1–12, ARTKQTARKSTG) at a 1:10 molar ratio before crystallization. Crystals were grown by the hanging-drop, vapour-diffusion method by mixing 1 μl protein (10 mg ml^{-1}) with 1 μl reservoir solution containing 100 mM sodium acetate (pH 5.3–5.6) and 600 mM ammonium sulphate at 18 °C. Crystals were cryoprotected by the reservoir buffer with 25% ethylene glycol and then flash frozen in liquid nitrogen.

The data were collected on beamline BL17U at Shanghai Synchrotron Radiation Facility, China, at wavelengths of 1.2816 Å and 0.9792 Å, respectively. Data were indexed, integrated, and scaled using the program HKL2000 (ref. 26). The orientation and position of CD–C^{DNMT3L} in the ADD–CD–C^{DNMT3L} complex was first determined by molecular replacement using CD–C^{DNMT3L} (PDB accession number 2QRV)[22] as a searching model in the PHASER program[27]. The resulting CD–C^{DNMT3L} model was refined with PHENIX package[28]. The ADD domain then was put into the refined extra-difference density using MOLREP with the 'search for model in the map' module[29]. The overall structure of the ADD–CD–C^{DNMT3L} complex was finally refined with stereochemistry and the reference structure ADD domain (PDB accession number 3A1B)[16] as restraints. The anomalous Fourier map of zinc cations in the ADD domain confirmed the validity of the position of the ADD domain. The structure of the ADD–CD–C^{DNMT3L}-H3 complex was determined by molecular replacement using the ADD domain (PDB accession number 3A1B)[16] and CD–C^{DNMT3L} (PDB accession number 2QRV)[22] as searching models in the PHASER program[27], and was then manually built by COOT[30].

All refinements used the module phenix.refine of PHENIX[28]. The model quality was checked with the PROCHECK program[31]. In the structure of ADD–CD–C^{DNMT3L}, 85.7% of residues were in most favoured regions, 12.9% in additional allowed regions, 0.9% in generously allowed regions, and 0.5% in disallowed regions. In the structure of ADD–CD–C^{DNMT3L}–H3, 88.4% of residues were in most favoured regions, 11% in additional allowed regions, and 0.6% in generously allowed regions. All structure figures were generated by PyMol[32].

***In vitro* DNA methylation assay on naked DNA.** The enzymatic activity of DNMT3A proteins was assessed by incorporation of a ^{3}H-labelled methyl group from *S*-adenosyl-L-[methyl-^{3}H]methionine ([methyl-^{3}H]AdoMet, PerkinElmer)[22,33]. A biotin-labelled DNA fragment amplified from the EBNA1 region of p220.2 (1.2 kilobases, 52 CG sites) was used as a substrate. For histone H3 stimulation, 0.3 μM DNMT3A proteins were pre-incubated with or without 3 μM histone H3 peptides (residues 1–12). For ADD-mediated inhibition, CD proteins (1 μM) were supplemented with or without the ADD domain proteins (18 μM) on ice for 30 min. DNA (100 ng) was methylated by DNMT3A proteins in the presence of 2.5 μM [methyl-^{3}H]AdoMet, 25 mM Tris–HCl (pH 7.5), 5% glycerol, 0.01% 2-mercaptoethanol, and 0.5 mg ml^{-1} BSA. The reactions were incubated at 37 °C for 30 min, and terminated by adding cold wash buffer (500 mM NaCl and 1 mM EDTA in PBST). The DNA products were immobilized on streptavidin beads, washed five times, and subjected to liquid-scintillation counting (PerkinElmer). Each reaction was performed in triplicate.

***In vitro* DNA methylation assay on recombinant poly-nucleosomes.** Recombinant *Xenopus laevis* histones were expressed and purified as described[34]. Site-specific methylation of $H3K_C4me3$ was performed by the methyl-lysine analogs approach[35]. Incorporation of $H3K_C4me3$ was verified by specific antibodies (anti-H3K4me3: Cell Signaling, 9751 s; anti-H3: Abcam, ab1791) and visualized on Tanon-5200 Chemiluminescent Imaging System (Tanon Science & Technology). Assembly of histone octamers and reconstitution of poly-nucleosomes were performed by salt dialysis using the 601 sequence (30 CG sites)[34]. The reaction mixture was the same as described above except that 300 ng nucleosomal DNA was methylated using 1 μM DNMT3A proteins. The reaction was quenched by the addition of excess TE buffer and 1% SDS. The histone proteins were removed by phenol–chloroform extraction. Then DNA was purified by ethanol precipitation, resolved in TE, and subjected to liquid-scintillation counting (PerkinElmer). Each reaction was performed in triplicate.

Electrophoretic mobility-shift assay. A 6-carboxy-fluorescein (FAM)-labelled double-stranded DNA (dsDNA) containing one CpG site was generated from annealing two primers (upper primer, FAM-5′-CTGAATACTACTTGCGCTCTCTAACCTGAT-3′; lower primer, 5′-GACTTATGATGAACGCGAGAGATTGGACTA-3′). The FAM-labelled DNA was used both in electrophoretic mobility-shift assays and fluorescence polarization assays. DNA (25 nM) and the indicated amounts of proteins were incubated in reaction buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 8% glycerol, and 0.5 mg ml^{-1} BSA for 30 min at 25 °C. The samples were subject to a 12% PAGE and analysed by Typhoon FLA 9500 (GE Healthcare) image scanning.

Fluorescence polarization assay. FAM-labelled dsDNA (15 nM) was incubated with increasing amounts of DNMT3A proteins for 30 min at 25 °C in reaction buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 8% glycerol, and 0.5 mg ml^{-1} BSA. Fluorescence polarization measurements were performed on a Synergy 4 Microplate Reader (BioTek) at 25 °C. The bound fractions were calculated as (mP − baseline mP)/(maximum mP − baseline mP), in which mP (milli-polarization units) represents the fluorescence polarization value. For the peptide stimulation experiment, 15 nM DNA was pre-incubated with 0.2 μM CD–C^{DNMT3L} or 1 μM ADD–CD–C^{DNMT3L} protein. An increasing amount of histone H3 peptide (H3K4me0 or H3K4me3) was then added into the protein–DNA complex. The levels of protein–DNA complex formation were measured by fluorescence polarization. Each reaction was performed in triplicate. The curves were fitted using GraphPad Prism 5.

Isothermal titration calorimetry. To obtain the binding affinity between ADD–CD–C^{DNMT3L} and H3 peptide, 0.04 mM ADD–CD–C^{DNMT3L} (in cell) in the absence or presence of 0.5 mM dsDNA (30 base pairs, upper strand: 5′-CTGAATACTACTTGCGCTCTCTAACCTGAT-3′) was titrated with 0.5 mM histone H3 peptide (residues 1–12) (in syringe) using an iTC200 microcalorimeter (GE Healthcare) at 18 °C. Protein, DNA, and peptide were prepared in a buffer containing 10 mM HEPES, pH 8.0, 100 mM NaCl, and 0.5 mM TCEP. To obtain the binding affinity between the ADD domain and H3 peptide, 0.05 mM ADD protein (in cell) was titrated with 0.5 mM histone H3 peptide (residues 1–12) (in syringe). The data were fitted by Origin 7.0 software.

***In vitro* binding assay.** For the GST pull-down assay, 30 μg DNMT3A-CD proteins were incubated with 8 μg GST–ADD–linker fusion proteins for 1 h at 4 °C in binding buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.01% 2-mercaptoethanol, 5% glycerol, and 0.1% Triton X-100. GST–ADD–linker proteins were then immobilized to 25 μl of glutathione resins (GE Healthcare) for 1 h at 4 °C. After washing three times with binding buffer, bound proteins were subjected to SDS–PAGE and stained by Coomassie blue.

For the histone peptide pull-down assay, 1 μg biotinylated histone H3 peptide (residues 1–21) was incubated with 30 μg wild-type and mutant ADD–CD–C^{DNMT3L} proteins for 1 h at 4 °C in binding buffer containing 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.01% 2-mercaptoethanol, 5% glycerol, and 0.1% Triton X-100. Then 20 μl streptavidin beads were added into the mixture and incubated 1 h at 4 °C. After washing three times with binding buffer, bound proteins were subjected to SDS–PAGE and stained by Coomassie blue.

Expression of ^{19}F-labelled proteins. An orthogonal tRNA/tRNA synthetase system was used to incorporate ^{19}F-labelled unnatural amino acid. Briefly, a TAG stop codon was introduced into the desired site to encode L-4-trifluoromethylphenylalanine (tfmF). A modified pEVOL-tfmFRS plasmid to express $tRNA_{CUA}$ and tfmF-specific aminoacyl-tRNA was co-transformed with the TAG-carrying plasmid into BL21 (DE3)[36]. The bacterial culture was induced with 0.02% L-arabinose, 1 mM tfmF, and 1 mM IPTG. The ^{19}F-labelled proteins were purified as for wild-type DNMT3A proteins.

^{19}F NMR spectra measurements. All one-dimensional ^{19}F NMR spectra measurements were performed at 293 K on an Agilent 500 MHz spectrometer equipped

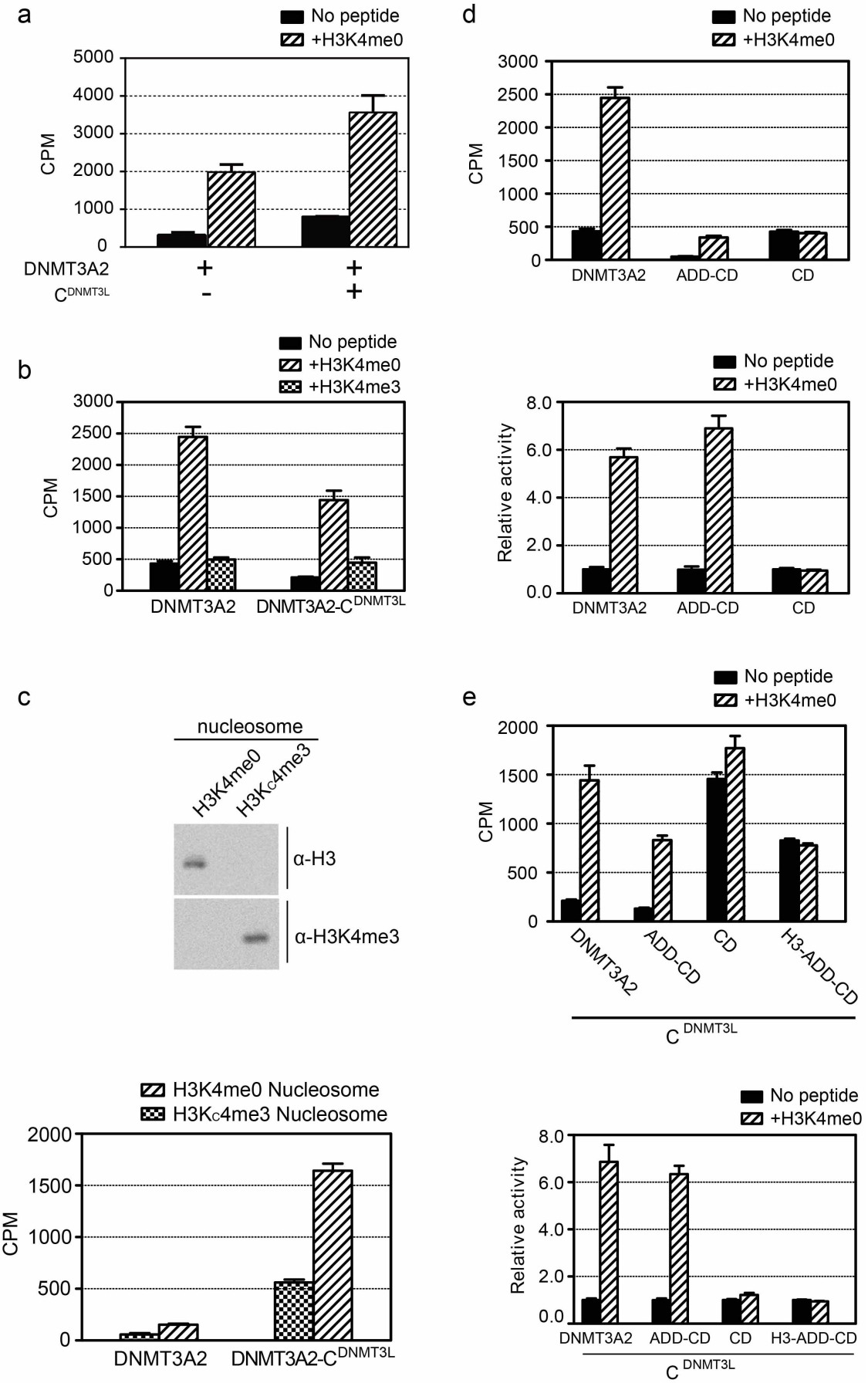
with an HFT probe, and the observation channel was tuned to ^{19}F (470.2 MHz), with 1,024 free induction decay accumulations in every 4 s recycling delay. A one-dimensional ^{19}F spectrum was acquired with one pulse program with a 90° pulse width of 12.45 μs and power at 57 W. The spectrum width was 60 p.p.m. and offset was −62 p.p.m. ^{19}F chemical shifts were referenced to an external standard, tfmF (−62.38 p.p.m.). The data were processed and plotted with an exponential window function (line broadening = 20 Hz) using ACD/NMR Processor Academic Edition software (ACD/Labs). The spectra of 0.14 mM ADD–CD F686tfmF or 0.12 mM ADD–CD F827tfmF with or without histone H3 peptides (residues 1–12) were collected at 293 K.

Electron microscopy data collection. DNMT3A2–C^{DNMT3L} in the absence or presence of histone H3 peptide was analysed by negative-stain electron microscopy in the same manner. The samples were prepared by dilution of purified protein complex to 6.15 μg ml^{-1}, then 4–5 μl of this sample was deposited onto a glow-discharged 400 mesh continuous carbon grid (Beijing Zhongjingkeyi Technology). The sample was then stained with 2% uranyl formate and air-dried. Data were recorded on a Tecnai G2 F20 TWIN transmission electron microscope (FEI) equipped with a field-emission gun operated at 200 kV. Images were recorded at ×71,000 microscope magnification on a 4k × 4k Eagle CCD (charge-coupled device) camera with a pixel size of 1.15 Å per pixel. The defocus ranged from −0.5 to −0.8 μm.

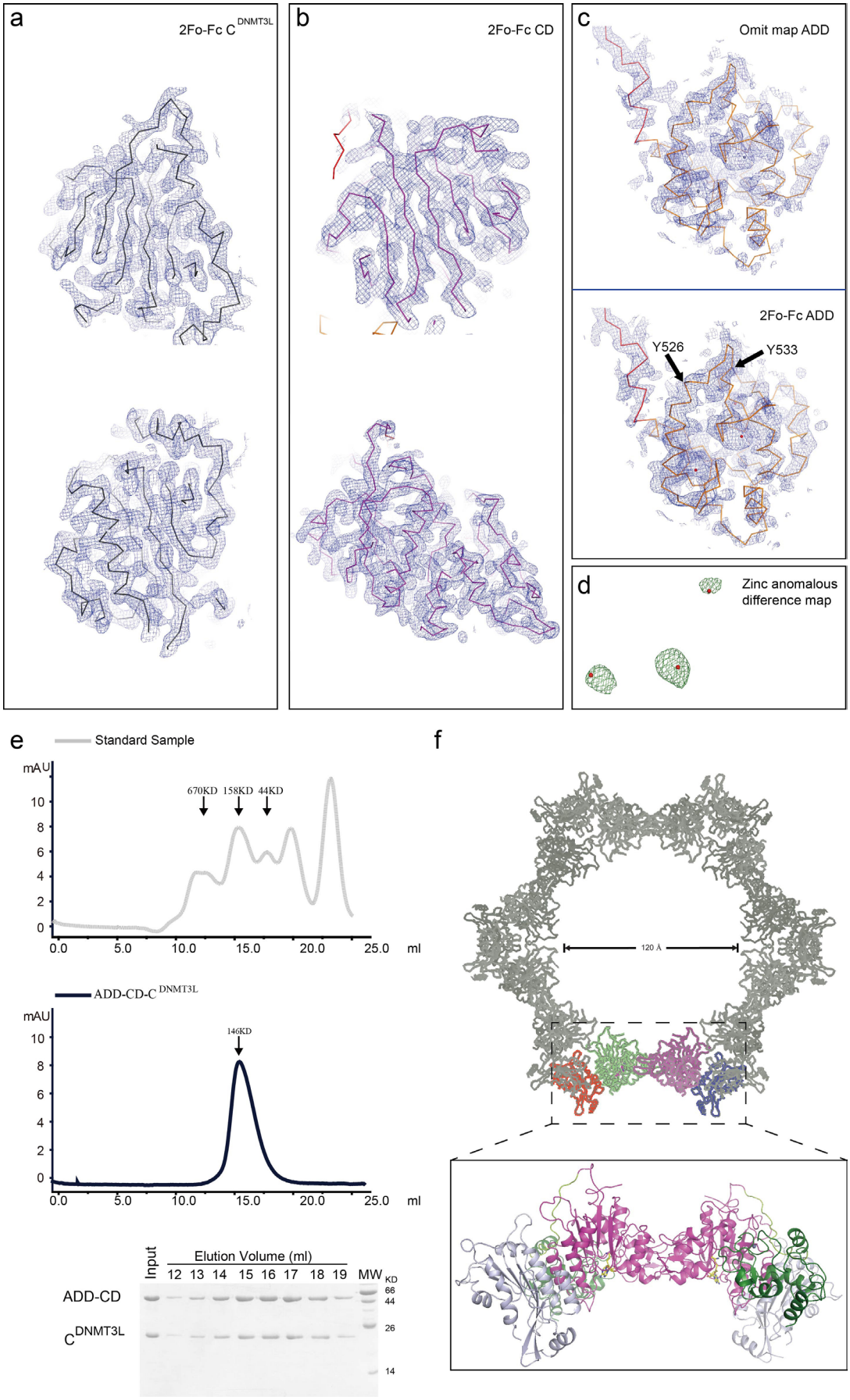
Electron microscopy image processing and three-dimensional reconstruction. For this, 9,864 and 21,910 particles were boxed out for DNMT3A2–C^{DNMT3L} and DNMT3A2–C^{DNMT3L}-H3, respectively, by using the e2boxer.py program in EMAN2.1 (ref. 37). Contrast transfer function parameters were determined for particles boxed out from each CCD image using EMAN1.9 procedure ctfit, followed by phase flipping using the applyctf program. The data were then low-pass filtered to 10 Å to enhance the image contrast for three-dimensional reconstruction[38,39]. Reference-free two-dimensional analysis used the EMAN1.9 program refine2d.py and IMAGIC[40], and those class-averages were used to generate initial models by e2initialmodel.py in EMAN2.1. Three-dimensional reconstruction was performed by the EMAN1.9 program refine[41,42]. Initially, no symmetry was imposed in the reconstruction process, and the resulting three-dimensional reconstruction revealed the existence of a two-fold symmetry in both maps but in different locations, which was subsequently imposed in the reconstruction process. The final resolution was estimated at 24 Å and 20 Å, respectively, by the 0.5 FSC criteria using the eotest program in EMAN1.9.

UCSF Chimera (http://www.cgl.ucsf.edu/chimera/)[43] was used to render the electron microscopy density together with the crystal structures. In addition, the Fit In Map module in Chimera was used for rigid-body fitting of the crystal structure into the corresponding electron microscopy density map.

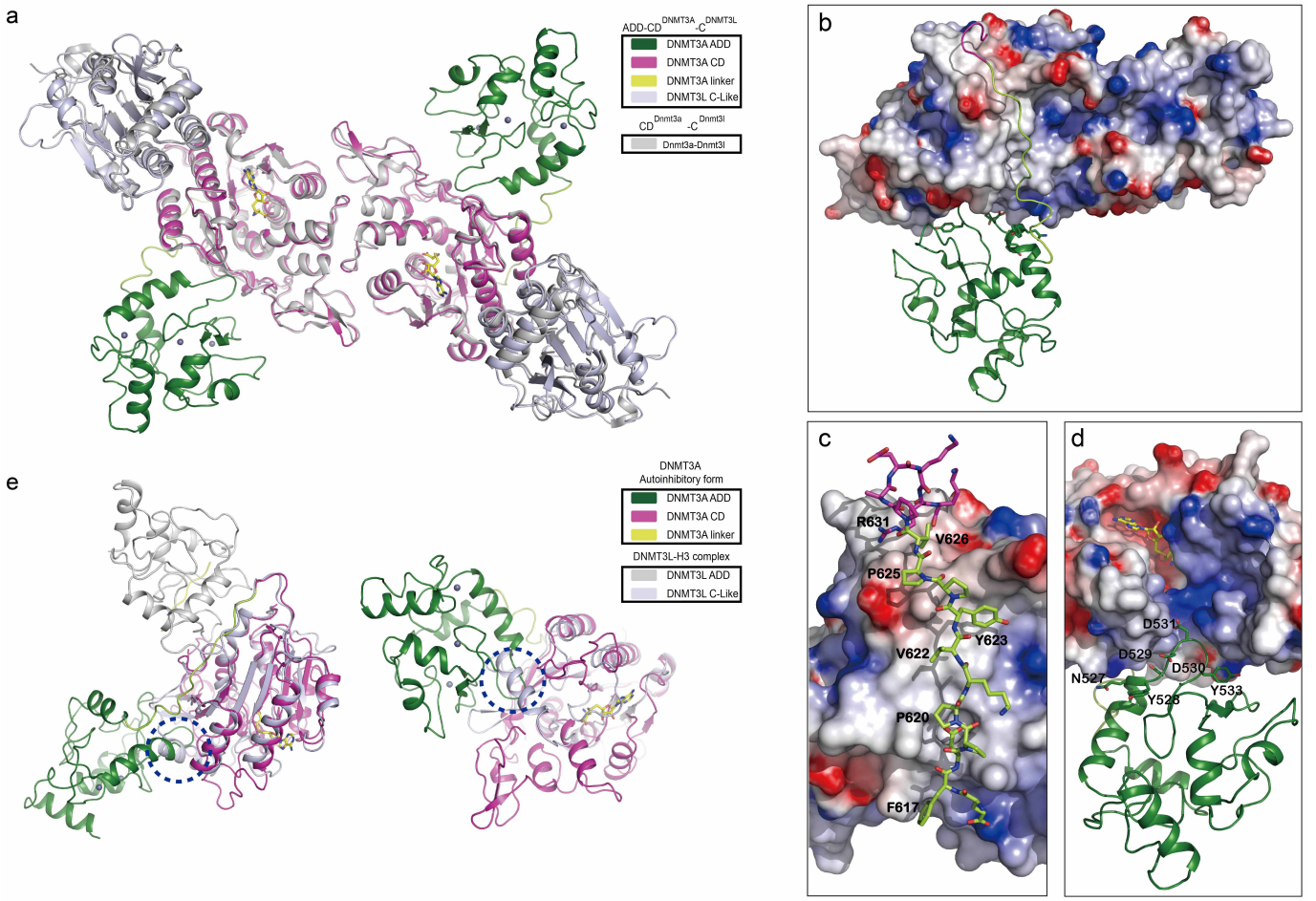
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
27. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
28. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
29. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022–1025 (1997).
30. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–291 (1993).
32. DeLano, W. L. The PyMOL Molecular Graphics System. http://www.pymol.org (2002).
33. Roth, M. & Jeltsch, A. Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biol. Chem.* **381,** 269–272 (2000).
34. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119,** 1–16 (1999).
35. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003–1012 (2007).
36. Shi, P. *et al.* Site-specific (1)(9)F NMR chemical shift and side chain relaxation analysis of a membrane protein labeled with an unnatural amino acid. *Protein Sci.* **20,** 224–228 (2011).
37. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
38. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
39. Ludtke, S. J., Jakana, J., Song, J. L., Chuang, D. T. & Chiu, W. A 11.5 Å single particle reconstruction of GroEL using EMAN. *JMB* (2001).
40. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
41. Cong, Y. & Ludtke, S. J. Single particle analysis at high resolution. *Methods Enzymol.* **482,** 211–235 (2010).
42. Cong, Y. *et al.* 4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl Acad. Sci. USA* **107,** 4967–4972 (2010).
43. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).



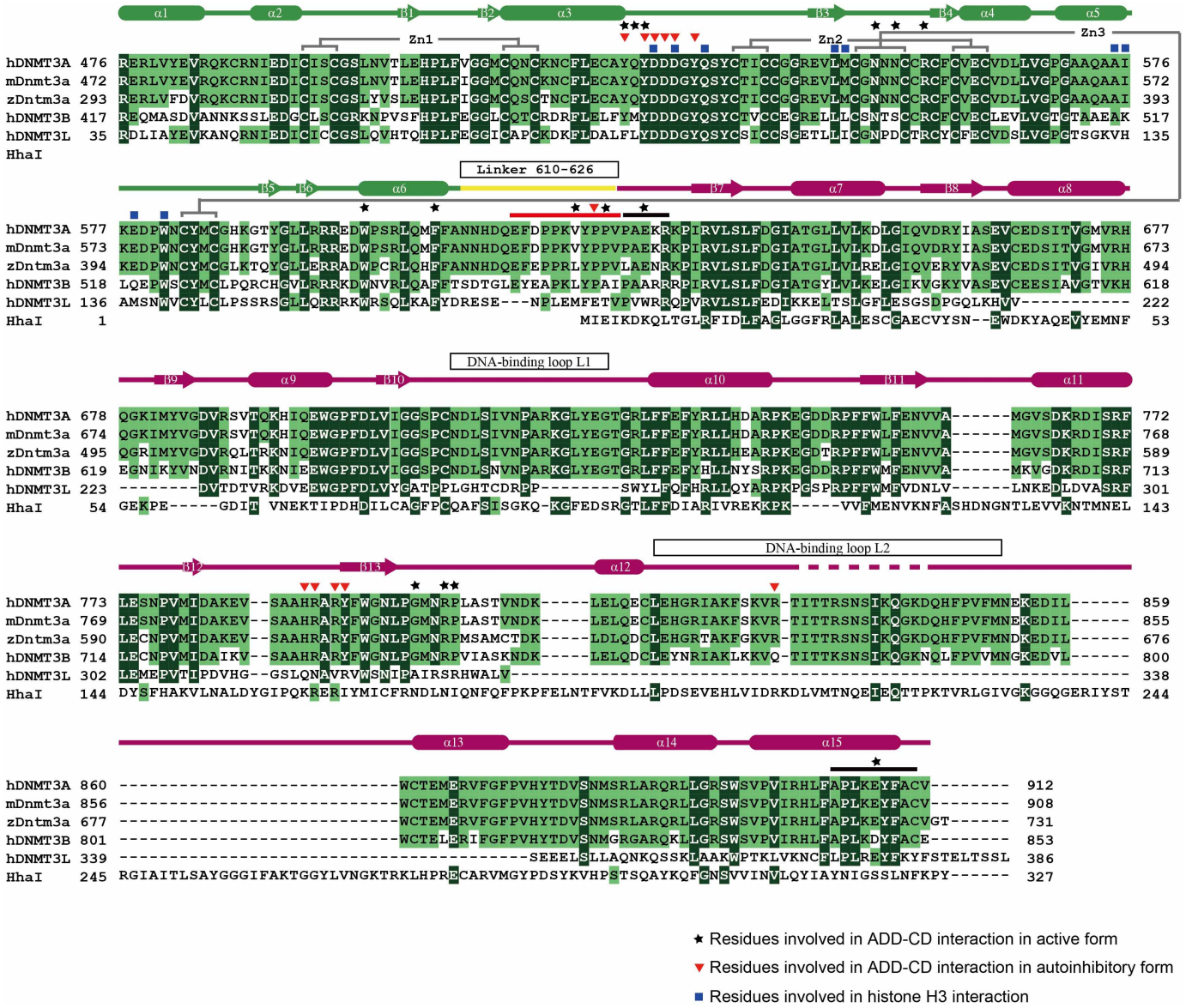
Extended Data Figure 1 | *In vitro* DNA methyltransferase activity of DNMT3A. **a**, *In vitro* DNA methyltransferase activities of DNMT3A2 (purified from insect cells) in the absence or presence of C^{DNMT3L}. The assays were performed in the presence or absence of histone H3 peptide (residues 1–12). Note that C^{DNMT3L} could enhance the activity of DNMT3A2 by a factor of 2–3, which is consistent with previous study[22]. However, histone H3-mediated activation of DNMT3A is independent of the existence of C^{DNMT3L}. **b**, *In vitro* DNA methyltransferase activities of DNMT3A2 (purified from insect cells) or DNMT3A2–C^{DNMT3L} (purified from bacteria) in the presence or absence of histone H3 peptides. **c**, Enzymatic activities of DNMT3A2 (purified from insect cells) or DNMT3A2–C^{DNMT3L} (purified from bacteria) using reconstituted nucleosomes as substrates. Nucleosomes containing unmodified histone H3 or $H3K_C4me3$ were subject to SDS–PAGE and visualized using specific antibodies. **d**, **e**, Enzymatic activities of various N-terminal deletions of DNMT3A2 in the absence (**d**) or presence (**e**) of C^{DNMT3L}. Corresponding relative activities are indicated at the bottom of each figure. CPM, counts per minute. Error bars, s.d. for triplicate experiments. The ADD–CD or CD protein purified from bacteria was not stable in solution and tended to precipitate out, which may have resulted in their lower activities under our experimental conditions (compared with DNMT3A2 purified from insect cells). Because C^{DNMT3L} could stabilize DNMT3A and had no effect on histone H3-mediated activation, protein complexes ADD–CD–C^{DNMT3L} and CD–C^{DNMT3L} were used in the following studies if not specified.



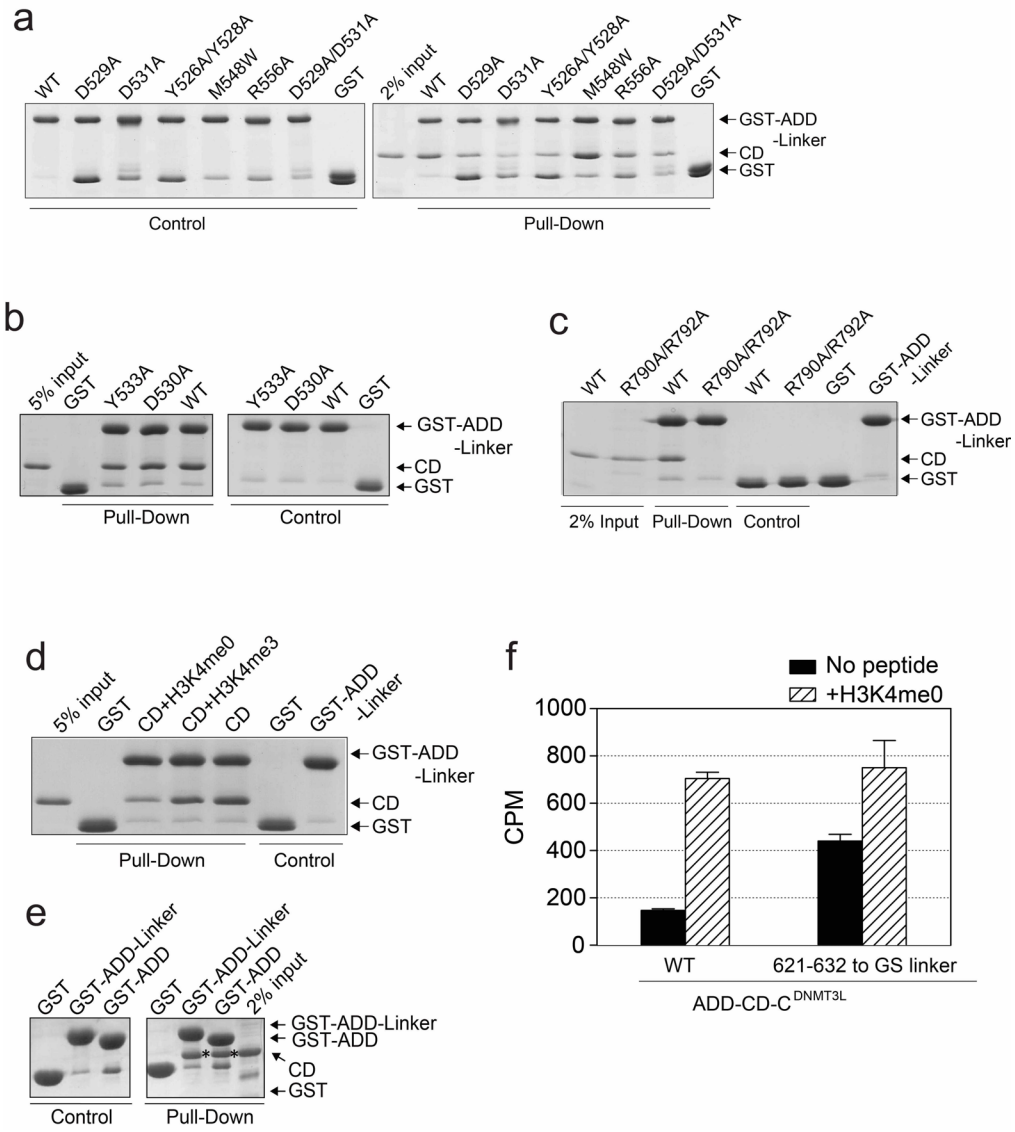
Extended Data Figure 2 | Crystal structure of ADD–CD–C^{DNMT3L} in autoinhibitory form. **a**, **b**, Two different views of the $2F_{observed} - F_{calculated}$ maps for C^{DNMT3L} (**a**) and CD (**b**) domains in the ADD–CD–C^{DNMT3L} structure. The maps were calculated at 3.82 Å and contoured at 1.5σ. Only main-chains are shown for simplicity. **c**, The $2F_{observed} - F_{calculated}$ maps for the ADD domain after refinement of the CD–C^{DNMT3L} complex (top) and after refinement of the ADD–CD–C^{DNMT3L} complex (bottom). The maps were calculated at 3.82 Å and contoured at 0.8σ. Main-chains from most residues, including residues 526–533 involved in the interaction with CD domain, fit well into the electron density. Some loop regions were not well covered by electron density, which is consistent with a high B factor (Extended Data Fig. 8a) of the ADD domain in the complex structure, supporting the dynamic feature of the ADD domain for regulating enzymatic activity of DNMT3A. **d**, Zn-anomalous difference map contoured at 3.5σ shows the positions of zinc cations in the ADD domain. **e**, Gel filtration profiles for standard proteins and the ADD–CD–C^{DNMT3L} complex. The peak position corresponds to the dimer of ADD–CD–C^{DNMT3L} with a molecular mass of about 140 kDa. **f**, Dimer formation of the ADD–CD–C^{DNMT3L} complex in crystals. The dimer of ADD–CD–C^{DNMT3L} complexes is mediated by CD–CD interaction in a two-fold crystallographic symmetry. Given the difficulty in tracing the conformation of the side chain in 3.82 Å resolution structure, we have not discussed the specific hydrogen bond or hydrophobic interaction within ADD–CD–C^{DNMT3L}. Residues 832–846 of DNMT3A were not built in the model because they lacked electron density, which may have resulted from their flexibility in crystals.



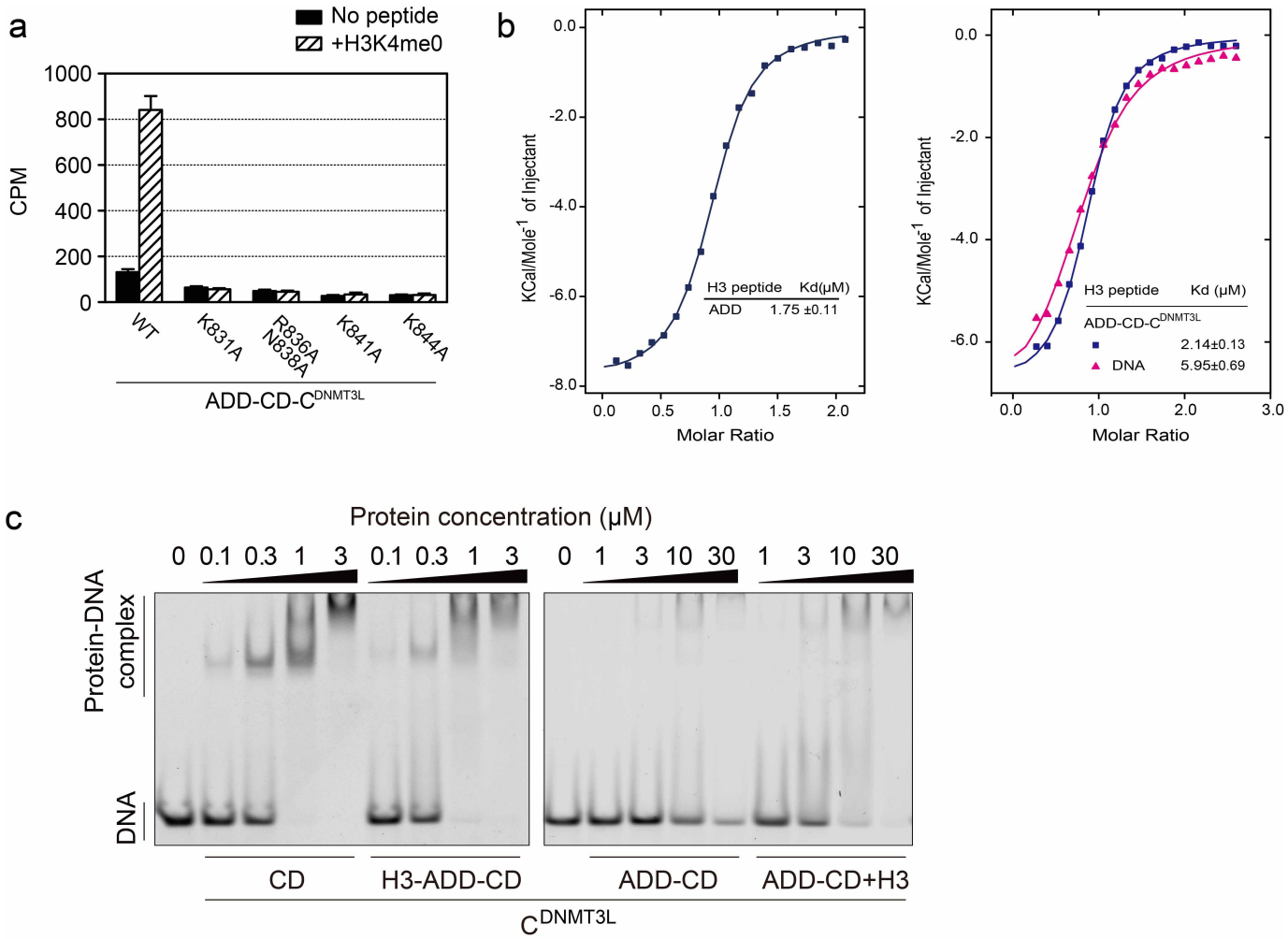
Extended Data Figure 3 | Structure of ADD–CD–C^{DNMT3L} in autoinhibitory form. **a**, Superimposition of human ADD–CD–C^{DNMT3L} with mouse CD^{Dnmt3a}-C^{Dnmt3L} (lack of ADD domain, PDB accession number 2QRV)[22] structures shown in ribbon representations. CD–C^{DNMT3L} in two structures is well aligned with a root mean squared deviation of 1.28 Å for 723 Cα aligned. The function of DNMT3A–DNMT3L complex dimerization has been characterized in a previous study[22]. The functions and structures of the CD and C^{DNMT3L} domains, and the CD–CD and CD–C^{DNMT3L} interfaces, were not discussed in this work. **b**, Overall structure of ADD–CD–C^{DNMT3L} with CD–C^{DNMT3L} shown in electrostatic potential surface, and the ADD domain and linker shown in ribbon representation. The linker packs against a hydrophobic surface of the CD domain. **c**, **d**, Close-up view of linker-CD (**c**) and ADD–CD (**d**) interfaces with the electrostatic potential surface of the CD domain indicated. Critical residues are shown in stick representation. **e**, Superimposition of ADD–CD–C^{DNMT3L} and DNMT3L-H3 structures (PDB accession number 2PVC)[15] shown in ribbon representations in two different views. The CD domain and C-like domain of DNMT3L were aligned for comparison. Note that the extended loop of the ADD domain in ADD–CD–C^{DNMT3L} overlaps with an α helix in the DNMT3L–H3 structure. DNMT3L is unlikely to adopt a similar conformation to that of ADD–CD–C^{DNMT3L} because otherwise the ADD domain will have steric hindrance with the C-like domain of DNMT3L (dashed circle). According to the above analyses, the structure of the autoinhibitory form of DNMT3A could not be predicted on the basis of the DNMT3L structure because the overall structures of DNMT3A and DNMT3L are different.



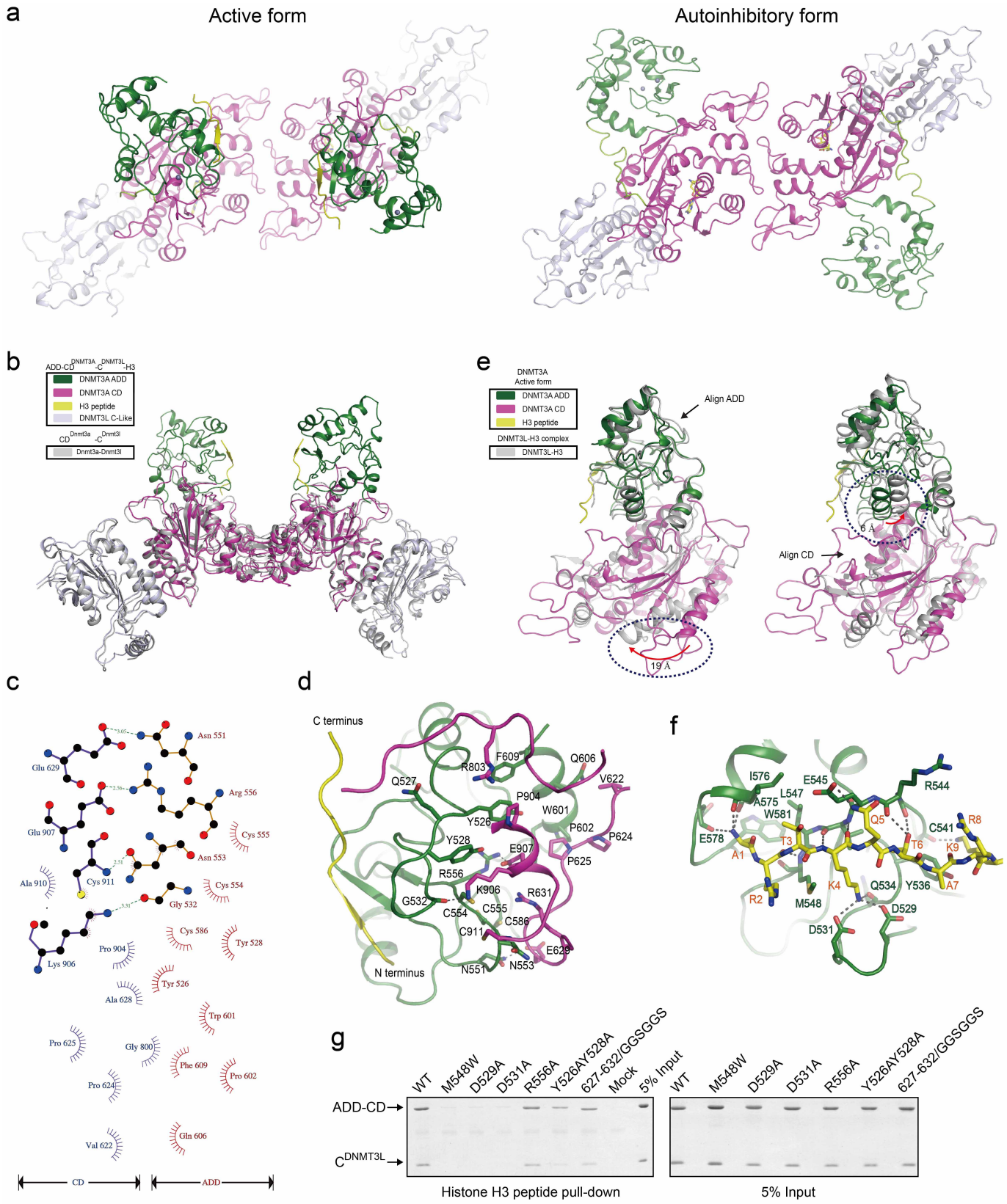
Extended Data Figure 4 | Sequence alignment of DNMT3 family members. Sequences of human DNMT3A (NP_072046), DNMT3B (NP_008823), DNMT3L (NP_787063), mouse Dnmt3a (NP_031898), zebrafish Dnmt3a (NP_001018150), and DNA methyltransferase from *Haemophilus parahaemolyticus* (WP_005706946) used in the alignment. Highly conserved and identical residues are highlighted with dark green background, and conserved residues are indicated with light green background. Secondary structural elements are coloured as in Fig. 1a and indicated above the sequences. Invisible residues in the structure of ADD–CD–C^{DNMT3L} are indicated as dashed lines above the sequences. Residues involved in ADD–CD interactions in active form or autoinhibitory form are indicated as black stars and red triangles, respectively. Residues involved in H3–ADD interactions are indicated as blue squares.



Extended Data Figure 5 | Interactions between the ADD and CD domains. **a**, **b**, GST pull-down assays with recombinant CD (residues 627–912) protein incubated with wild-type or mutant GST–ADD–linker proteins immobilized on glutathione resin. The bound proteins were analysed by SDS–PAGE and Coomassie blue staining. **c**, GST pull-down assays using wild-type or mutant of the CD domain. **d**, GST pull-down assays in the absence or presence of histone H3 peptide (H3K4me0 or H3K4me3). **e**, GST pull-down assays with the CD domain incubated with GST–ADD or GST–ADD–linker proteins immobilized on glutathione resin. **f**, Activities of wild-type and mutant ADD–CD. Residues 621–632 were replaced by a GS linker in the mutant proteins.



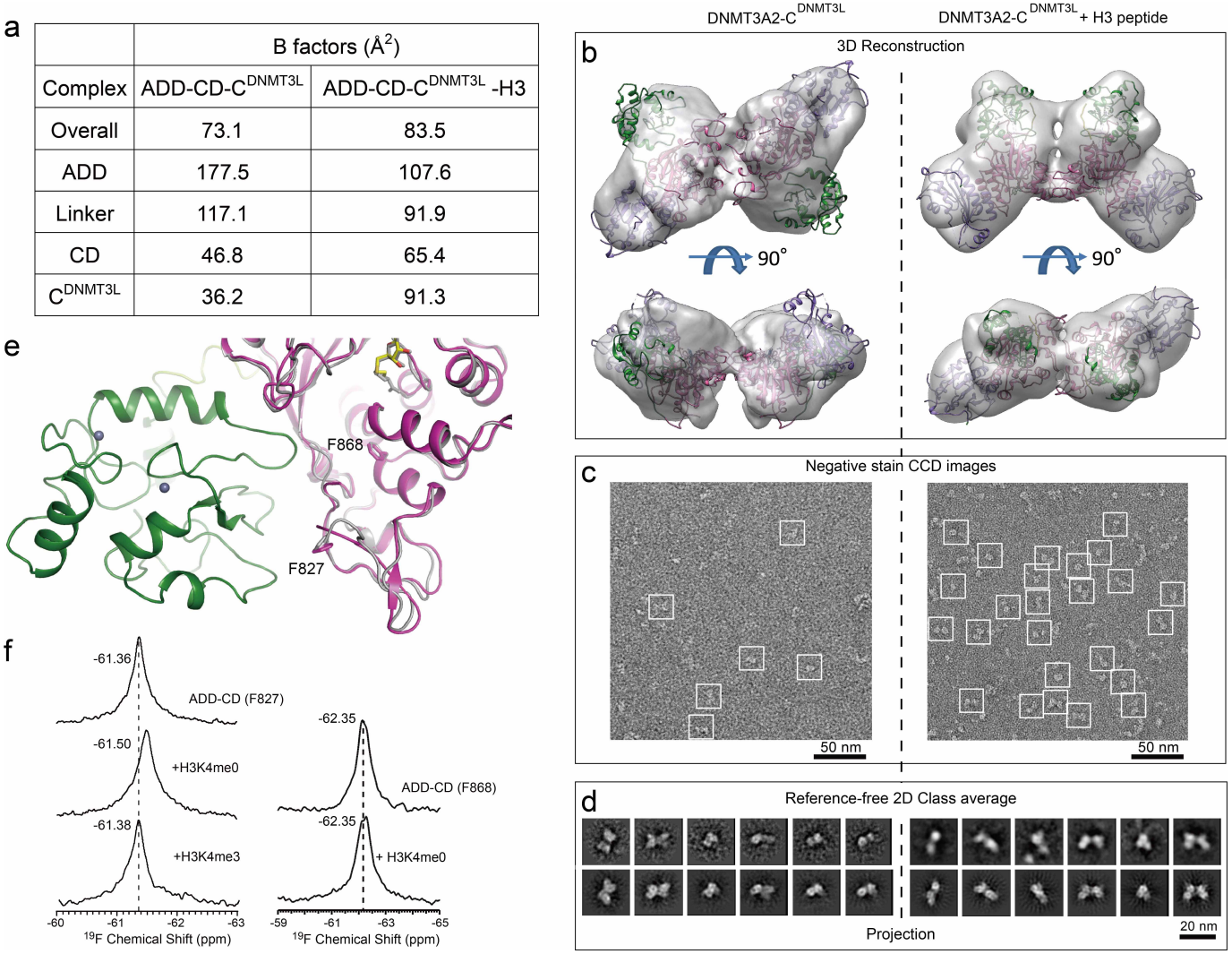
Extended Data Figure 6 | Interactions between DNMT3A and DNA. **a**, Enzymatic activities of wild-type and mutant ADD–CD–C^{DNMT3L}. Residues on the missing loop (residues 831–846) were mutated for the *in vitro* DNA methyltransferase activity assay. Error bars, s.d. for triplicate experiments. Mutating above residues leads to loss of activity of ADD–CD–C^{DNMT3L}, supporting their important role in catalysis or DNA recognition. The missing loop in the ADD–CD–C^{DNMT3L} structure is equivalent to a DNA-binding loop in the HhaI–DNA structure. **b**, DNA has no effect on the interaction between histone H3 and DNMT3A. Left, isothermal titration calorimetry enthalpy plot for the binding of isolated ADD domain (in cell) to histone H3 peptide (residues 1–12, in syringe), with the estimated binding affinities (K_d) listed. Right, superimposed isothermal titration calorimetry enthalpy plots for the binding of ADD–CD–C^{DNMT3L} (in cell) to histone H3 peptide (residues 1–12, in syringe) in the absence or presence of dsDNA. The estimated binding affinities (K_d) are listed. Histone H3 peptide has comparable binding affinity to the ADD domain alone (1.75 μM) and ADD–CD–C^{DNMT3L} in autoinhibitory form (2.14 μM), and the addition of DNA was not able to enhance the binding affinity further. The presence of DNA led to a slight decrease in the binding affinity between histone H3 peptide and ADD–CD–C^{DNMT3L}, which may have resulted from slight precipitation of the protein caused by the high concentration of DNA used for titration. **c**, Electrophoretic mobility-shift assays for DNMT3A proteins in the absence or presence of histone H3 peptide, with protein concentrations indicated. H3–ADD–CD represents a fusion protein with histone H3 (residues 1–20) at the N terminus of ADD–CD. The assays showed that CD–C^{DNMT3L} strongly bound to the FAM-labelled DNA duplex, whereas the existence of the ADD domain markedly decreased DNA-binding affinity, which was partly restored by the addition of histone H3 peptide or largely restored by H3–ADD–CD fusion protein.



Extended Data Figure 7 | Structure of ADD–CD–C^{DNMT3L}-H3 in active form. **a**, Ribbon representations of the overall structures of the ADD–CD–C^{DNMT3L} in active (left) and autoinhibitory (right) forms. Histone H3 peptides are coloured in yellow. **b**, Structural comparison of human ADD–CD–C^{DNMT3L}–H3 and mouse CD^{Dnmt3a}–C^{Dnmt3L} complexes. The compared structures are shown in ribbon representations. The ADD–CD–C^{DNMT3L} complex structure (this study) is coloured as in Fig. 1d, and the CD^{Dnmt3a}-C^{Dnmt3L} complex structure[22] is coloured in grey. Residues 611–620 and 833–846 of DNMT3A were not built in the model because they lacked electron density. **c**, LIGPLOT representation of the ADD–CD interactions in the ADD–CD–C^{DNMT3L}-H3 structure. Carbon, oxygen, and nitrogen are shown as black, red, and blue balls, respectively. Hydrogen bonds are indicated as dashed lines, with lengths given in Å. **d**, Close-up view of the ADD–CD interface. Critical residues for the interactions are shown in stick representation, and hydrogen bonds are indicated as dashed lines. The C terminus (residues 903–911) of the CD domain and a loop region (residues 621–632) together form a flat patch for interaction with the ADD domain. Hydrogen bonds are formed between residues N551, N553, and R556 of the ADD domain and residues E629, C911, and E907 of the CD domain. Residues Y526, Y528, W601, and F609 of the ADD domain, V622 and P625 of the linker, and R803 and P904 of the CD domain are involved in hydrophobic interactions. **e**, Structural comparison of ADD–CD–H3 in ADD–CD–C^{DNMT3L}–H3 (this study) and DNMT3L–H3 structures (PDB accession number 2PVC)[15]. Two compared structures are shown in ribbon representations with ADD domains (left) or catalytic domains (right) aligned, respectively. The DNMT3L–H3 structure is coloured in grey. When the ADD domains are superimposed, the catalytic domain moves with a longest distance of 19 Å. When the CD domains are superimposed, the ADD domain moves 6 Å. **f**, Close-up view of the H3–ADD interface. Critical residues for the interactions are shown in stick representation, and hydrogen bonds are indicated as dashed lines. The fashion of histone H3–ADD interaction is similar to that observed in the structure of the H3–ADD fusion protein[16]. **g**, Histone H3 peptide pull-down assay. Recombinant wild-type and mutant ADD–CD–C^{DNMT3L} proteins were incubated with biotinylated histone H3 peptide (residues 1–21) and immobilized onto streptavidin sepharose beads. Bound proteins were subjected to SDS–PAGE and stained by Coomassie blue.

a

	B factors (Å^2)	
Complex	ADD-CD-C^{DNMT3L}	ADD-CD-C^{DNMT3L} -H3
Overall	73.1	83.5
ADD	177.5	107.6
Linker	117.1	91.9
CD	46.8	65.4
C^{DNMT3L}	36.2	91.3



Extended Data Figure 8 | Conformational change of DNMT3A induced by histone H3 tail. **a**, Average B factors for domains of ADD–CD–C^{DNMT3L} in the structures of ADD–CD–C^{DNMT3L} and ADD–CD–C^{DNMT3L} bound to H3 peptide. The average B factor of the ADD domain is higher than other domains in both structures, and is higher in autoinhibitory form (177.5 Å^2) than that in active form (107.6 Å^2). The results indicate that the ADD domain is more dynamic than other domains of the complex, especially in its autoinhibitory form. The observation further supports the idea that DNMT3A undergoes conformational changes on the ADD domain induced by histone H3. **b**, Two different views of the electron microscopy density maps of DNMT3A2–C^{DNMT3L} (left) and DNMT3A2–C^{DNMT3L}-H3 (right) processed to 24 Å and 20 Å resolution, respectively. The corresponding crystal structure was fitted into the electron microscopy density map for each state. The density is not fully occupied, which might because of the missing PWWP domain in the crystal structures. **c**, Typical negative stain CCD images of DNMT3A2–C^{DNMT3L} (left) and DNMT3A2–C^{DNMT3L}-H3 (right). Representative particles are highlighted by white boxes. **d**, Comparison of the two-dimensional projections (bottom) from the electron microscopy map with the corresponding reference-free two-dimensional class averages (top) reveals similar structural features. **e**, Position of residues F827 and F868 for ^{19}F NMR measurements. Close-up view of the DNMT3A structure in autoinhibitory form with residues F827 and F868 indicated in stick representation. Residue F827 is located in loop L2 (for DNA binding) and close to the ADD domain. As a negative control, residue F868 is located close to the catalytic cavity and away from the ADD domain. Residue F868 is unlikely to undergo conformational change when the ADD domain dissociates from the CD domain. To detect conformational changes of DNMT3A in solution, residues F827 and F868 were substituted by ^{19}F-labelled L-4-trifluoromethylphenylalanine (^{19}F-tfmF) in ADD–CD. **f**, One-dimensional ^{19}F NMR measurements were performed using ADD–CD with substitution of F827tfmF (left) or F868tfmF (right) in the absence or presence of H3K4me0 or H3K4me3 peptide. The chemical shift for each measurement is indicated.

Extended Data Table 1 | Data collection and refinement

	ADD-CD-C^{DNMT3L} complex	ADD-CD-C^{DNMT3L} complex with H3 peptide
Data collection		
Space group	P 6_3 2 2	P 6_5
Cell dimensions		
a, *b*, *c* (Å)	252.0, 252.0, 75.3	183.8, 183.8, 123.3
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	50.0 – 3.82 (3.96 – 3.82)*	50.0 – 2.90 (3.00 – 2.90)
R_{sym} or R_{merge}	0.098 (0.870)	0.095 (0.721)
$I / \sigma I$	13.2 (1.9)	20.5 (3.5)
Completeness (%)	99.9 (99.9)	99.0 (100.0)
Redundancy	5.7 (5.8)	9.9 (9.8)
Refinement		
Resolution (Å)	50.0 – 3.82 (3.96 – 3.82)*	50.0 – 2.90 (3.00 – 2.90)
No. reflections	20564	52407
R_{work}/ R_{free}	0.230 / 0.273	0.223/ 0.261
No. atoms		
Protein	5043	9820
H3 peptide		158
Ligand/ion	29	58
Water		11
B-factors (Å^2)		
Protein	73.2	83.2
H3 peptide		108.7
Ligand/ion	55.1	61.1
Water		60.5
R.m.s deviations		
Bond lengths (Å)	0.005	0.010
Bond angles (°)	0.92	1.40

* Highest resolution shell is shown in parentheses.

Extended Data Table 2 | Effect of DNMT3A mutants on autoinhibition

Mutations	Description	Effect
D529A	H3K4me0 recognition (residue K4) ADD-CD interface (autoinhibitory form)	Release
D531A	H3K4me0 recognition (residue K4) ADD-CD interface (autoinhibitory form)	Release
M548W	H3K4me0 recognition (residue A1)	Inhibition
Y526A/Y528A	ADD-CD interface (autoinhibitory form)	Release
Residues 627-632 replaced by GGSGGS	ADD-CD interface (in both forms)	Release
E907A	ADD-CD interface (active form)	No-change
R742G/P743G/K744S	Potentially for DNA recognition	No-change
K812A	Potentially for DNA recognition	No-change
R720G/K721S	Potentially for DNA recognition	Decrease
R836A/N838A	Potentially for DNA recognition	Decrease
K831A	Potentially for DNA recognition	Decrease
K841A	Potentially for DNA recognition	Decrease
K844A	Potentially for DNA recognition	Decrease
Residues 809-813 replaced by GGSGG	Potentially for DNA recognition	Decrease
Residues 821-846 replaced by GGSGGSGG	Potentially for DNA recognition	Decrease

All mutants were made on ADD–CD (residues 476–912) and the protein complexes ADD–CD–C^{DNMT3L} were used for the assays. 'Release' represents release of the autoinhibition by at least twofold activity enhancement. 'Inhibition' represents inhibition of activity and no response to H3 peptides. 'No-change' indicates that the mutants behaved similar to wild-type protein. 'Decrease' represents decrease or loss of enzymatic activity.